Malaysian Journal of Medicine and Health Sciences (ISSN 1675-8544); Vol 12 (1) January 2016

Original Article

Subtractive Hybridization Identifies Stem Cell-Associated Genes in an Acute Myeloid Leukemia with Poor Prognosis

Ngiow Shin Foong,[1] *Maha Abdullah,[1,2] Jasmine Lim,[1] Cheong Soon-Keng,[3] Seow Heng-Fong[1]

[1] Department of Pathology, Faculty of Medicine and Health Sciences,
Universiti Putra Malaysia, 43400 UPM Serdang, Malaysia.
[2]Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM Serdang, Malaysia.
[3]Hematology Unit, Department of Pathology, Universiti Kebangsaan Malaysia Medical Center,
56000 Cheras, Kuala Lumpur, Malaysia.

ABSTRACT

Introduction: Current prognostic markers have improved survival prediction, however, it has not advanced treatment strategies. Gene expression profiling may identify biological markers suitable as therapeutic targets. Leukaemia stem cell is associated with adverse outcome, however, its biological characteristics are still being investigated. We observed higher *in vitro* cell viability in acute myeloid leukaemia (AML) samples with poor prognosis, which may be stem cell related. **Objective:** The objective of this study was to profile highly expressed genes in an AML sample of poor prognosis/high viability and compare with a sample of good prognosis/low viability. **Method:** Subtractive hybridization was performed on two AML samples with high blast counts (>80%), a poor prognosis, PP (disease free survival, DFS<12 months) and a good prognosis, GP (DFS>12 months) sample. The PP sample had higher CD34+ counts (73% vs 46%) and higher cell viability than the GP sample. cDNA libraries were subsequently cloned and sequenced. **Results:** cDNA subtracted from the PP samples was identified as genes active during fetal/embryonic development (*LCOR, CNOT1, ORMDL1), HOX*- related genes (*HOXA3, PBX3, SF3B1*), hematopoiesis (*SELL, IL-3RA*) and aerobic glycolysis/hypoxia (*PGK1, HIGD1A*) -associated genes. Majority of GP clones isolated contained genes involved in oxidative phosphorylation, OXPHOS (*COXs, ATPs, MTND4* and *MTRNR2),* protein synthesis (including ribosomal proteins, initiating and elongation factors), chromatin remodeling (*H2AFZ, PTMA*), cell motility (*MALAT1, CALM2, TMSB4X*), and mitochondria (*HSPA9, MPO*) genes. **Conclusion:** Thus, the PP sample exhibited stem cell-like features while the GP sample showed cells at a high level of cell activity. These genes are potential prognostic markers and targets for therapy.

Keywords: Acute myeloid leukemia, Gene profiling, Subtractive hybridization, Leukemia stem cell

INTRODUCTION

Acute myeloid leukemia (AML) is a hematological malignancy characterized by 20% or more blasts in the peripheral blood or bone marrow. Median age of presentation is in the late 60s. Early death (death within 30 days from diagnosis) occurs in 19% of patients. Standard chemotherapy with anthracycline and arabinoside ('3 plus 7') achieves complete remission in 65% *de novo* AML. Median overall survival was 1119 days for patients aged 16-55 year old (N=554), 350 days for those aged 56-65 year old (N=437) and 80 days for patients aged 76-89 year old (N=968) (1). Thus, survival rates in patients remain poor. Identifying patients with poor response to treatment is important in tailoring best therapy. Current prognostication utilize genomic markers such as cytogenetic abnormalities as well as gene mutations, however, accurate prediction remains unsatisfactory (2-4). Gene expression profiling may reveal patterns that reflect underlying biology of disease subtypes (3).

AML is a heterogeneous disease, diverse in phenotype as well as genotype. Malignant transformation occur at early stages of hematopoietic development and differentiation which remains in WHO classification (5). Gene expression profiling has identified genes (4) and stem cell related genes (2) associated with poor prognosis in AML. There is limited literature on gene profiling of leukemia stem cells in AML. We observed cell viability of AML cells cultured for 3 days *in vitro* inversely correlated with disease free survival (6). High *in vitro* cell proliferation reflects self-renewal capacity associated with stem cells (7). Thus, AML of poor prognosis/high cell viability may be a source for leukaemia stem cells and provide an opportunity to dissect its biological properties.

*Corresponding Author: Assoc. Prof. Dr. Maha Abdullah
maha@upm.edu.my

The aim of this study was to identify differentially expressed genes in two AML samples from patients with different survival rates using subtractive hybridization. From the genetic profiles obtained it was shown that both samples exhibited distinct patterns in function and biology.

MATERIAL AND METHODS

Samples

Excess bone marrow/peripheral blood samples from acute myeloid leukemia patients were obtained from Hematology Laboratory, Universiti Kebangsaan Malaysia Medical Centre. Samples were collected between September 1999 and December 2002. Approval was obtained from the Institution and the Medical Research Ethics Committee, Universiti Putra Malaysia. All procedures were in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Treatment of AML patients were according to the Berlin-Frankfurt-Munich (BFM)-83 protocol with modifications. Disease free survival (DFS) was determined from the time complete remission was diagnosed till time first relapse was diagnosed or till end of study. Diagnosis was made by Hematologists. Poor prognosis was defined by short DFS (<12 months) and good prognosis defined by longer DFS (>12 months).

Two samples were selected based on poor prognosis (PP-AML) and good prognosis (GP-AML). PP-AML was a peripheral blood sample from a 46 year-old male, newly diagnosed AML, M4 FAB (French-American-British classification). Peripheral blood of PP-AML consists of 95% blast of which 73% were CD34+. GP-AML was a bone marrow sample from a 7 year-old female, newly diagnosed AML, M2 FAB. Peripheral blood of GP-AML consists of >99% blast. Bone marrow of GP-AML contained 46% CD34+ blast (and 95% CD13). Both samples were taken before commencement of chemotherapy. Selection for the two AML samples was also based on extreme ends in i*n vitro* cell viability determined in an earlier study(6); *in vitro* viability was high in poor prognosis (PP-AML) patient and low in good prognosis (GP-AML) patient. RNA from the peripheral blood of PP-AML and bone marrow of GP-AML were isolated for subtractive hybridization analysis.

Cell isolation and RNA extraction

Mononuclear cells from bone marrow / peripheral blood aspirates were isolated by gradient density centrifugation on Ficoll-Hypaque (Amersham Pharmacia Biotech AB, Uppsala, Sweden). Cells were stored under liquid nitrogen condition until further analysis.

Total RNA was extracted using TRI reagent based protocol (Tri-Reagent, Life Technology, USA). Quality and quantity determined on Nanodrop 1000 spectrophotometer (Thermo Scientific, Wilmington, USA) while RNA integrity was verified by detection of the 28S and 18S bands after electrophoresis on ethidium bromide stained agarose gels. Lambda DNA/EcoR I + Hind III DNA Marker (Promega, Wisconsin, USA) was run together on the gel. The lambda DNA is completely digested with EcoR I and Hind III and yields 12 double stranded Lambda DNA fragments: 21226, 5148, 4973, 4268, 3530, 2027, 1904, 1584, 1375, 947, 831, 564 bp.

cDNA synthesis and amplification

First strand cDNA was synthesized from RNA (1 mg) with the Super SMART™ PCR cDNA Synthesis Kit (Clontech, BD Biosciences) using the SMART CDS primer and SMART 11A oligonucleotide and PowerScript reverse transcriptase according to manufacturer's protocol. This was followed with column chromatography to purify resulting product. cDNA was then amplified by LD PCR with PCR Primer 11A and the Advantage 2 Polymerase Mix for amplification of long PCR templates, high yield and high fidelity (Advantage 2 PCR Kit, Clontech BD Biosciences). PCR was conducted in a thermal cycler (T-gradient Biometra, Montreal Biotech Inc., Kirkland, Canada) at thermal cycling profiles given in manufacturer's guide. A further purification step was then conducted with phenol:chloroform:isoamyl alcohol and n-butanol. As only amplified ds cDNA larger than 500 bp was to be used for further experiment, the resultant PCR product was run on standard agarose (1%), gel containing correct product size excised and purified with the QIAquick Gel Extraction kit (Qiagen, USA) according to manufacturer's manual. The Super SMART kit thus uses small amount of RNA and designed to enrich for full-length cDNA and ensure gene representation as well as amplify cDNA.

Subtractive hybridization and cloning

Following cDNA synthesis and amplification, procedure continued with the PCR-Select cDNA Subtractive Kit, (Clontech BD Biosciences, USA). *Rsa* I digestion was carried out for 3 hours at 37°C on both tester (AML sample 1) and driver (AML sample 2) to generate shorter, blunt-ended ds cDNA fragments. This was necessary to continue with adaptor ligation and subtraction. After this restriction enzyme digestion, the reaction was purified using Wizard SV Gel and PCR Cleanup System (Promega, USA) according to manufacturer's protocol. Subsequently, DNA was

precipitated with 4M ammounium acetate and 95% ethanol, pellet washed with 80% ethanol and finally dissolved in 6.7 microliter TNE buffer. cDNA before and after enzyme digestion was checked on gel electrophoresis. At this point one of the AML sample was chosen as the tester while the other as driver cDNA for the subsequent experiment. Adapter ligation was performed on tester cDNA (since this procedure will only enrich for differentially expressed sequences present in poly A+ RNA of the tester sample, this experiment was also performed with the cDNA samples reversed). Two adapter ligations were performed on each tester cDNA. Two aliquots were made of each cDNA; one was added with Adapter 1 while the other with Adapter 2R. After ligation with T4 DNA ligase, an aliquot was taken from each tube and mixed in a third tube to be the Unsubtracted tester control to act as the positive control for ligation and later as negative control for ligation. Before continuing with subtraction, the ligation efficiency analysis was conducted on ligated samples using PCR amplification method with primers supplied with the kit.

There are two steps in subtractive hybridization. In the first, excess driver cDNA was mixed with the two ligated tester cDNAs in separate tubes in hybridization buffer provided. The mixture was then heat denatured and allowed to anneal in a thermocycler (T-gradient Biometra, Montreal Biotech Inc., Kirkland, Canada) at 68^{o}C for 8 hr. Unhybridized ss cDNAs are the differentially expressed sequences enriched for the second hybridization step. In the second step, the two samples from first hybridization were mixed together and fresh denatured driver cDNA was added according to instructions in the manual. It is important not to denature products in the reactant tube at this point. Incubation was carried out at 68^{o}C overnight. The second incubation causes new hybrid molecules to form between differentially expressed cDNAs labeled with different adapters. The final step involved two PCR amplification, a primary PCR followed by nested PCR was performed on subtracted and unsubtracted cDNA and a PCR control subtracted cDNA provided in the kit. After combining all necessary reagents as detailed in the manual, the reaction was first incubated at 75^{o}C for 5 min to fill in missing strands of the adapters necessary as binding site for the PCR Primer 1 in the primary PCR. PCR Primer 1 binds the adapters. PCR amplification was carried out with the thermal cycling profile provided in the manual. Subsequently, 1 microliter from these reactions were used for nested PCR using nested primers provided in the kit. PCR reactions were amplified using the Advantage cDNA Polymerase Mix. Results were analyzed by gel electrophoresis. Thus, this kit selects for low-abundance mRNA as well as amplify and enrich for differentially expressed genes.

Subtracted libraries were cloned into pGEM-T vectors using pGEM-T Easy Vector system (Promega, USA) and transformed by heat shock method into JM109 Competent Cells (Promega, USA). All procedures were performed according to manufacturers' protocols. Recombinant strains were selected on LB-ampicillin (50 microgram/mL) agar plates spread with 50 microliter [1 of 5-bromo-4-chloro-3-indolyl-P-Dgalactopyranoside (40 mg/mL) (International Biotechnologies, Inc.) and eight microliter of 100 mM isopropyl-thiogalactoside (IPTG; Sigma Chemical Co., St. Louis, Mo.). White colonies were picked after 24 hours growth.

Colony polymerase chain reaction (PCR)

Colony PCR were performed to amplify inserts directly in *E. coli* colonies. Cell suspensions of selected colonies were heated at 94^{o}C for 10 min, followed by PCR in 0.5 microMolar of each primer pair, 0.2mM dNTP, 1 U of Taq DNA polymerase. *pGEM-T primers* T7F (5′-TAATACGACTCACTATAGGG-3′) and SP6R (5′-GATTTAGGTGACACTATAG-3′) were used. PCR conditions were as follows: 5 min at 94^{o}C for hot start, followed by 30 cycles of 94^{o}C for 40 sec, 56^{o}C for 40 sec for annealing, and 72^{o}C for 3 min, with a final extension of 4 minute at 72^{o}C . The amplified products were analyzed on 1% agarose gels and visualized by ethidium bromide staining.

Sequencing and Gene Identification

The amplified products were purified using Montage PCR plate kit (Millipore, US) according to manufacturer's protocol and DNA sequencing was performed by a commercial company (1st Base Laboratories Sendirian Berhad, Selangor, Malaysia). Briefly, the single pass DNA sequencing encompasses cycle sequencing reactions, dye terminator removal and analysis on Genetic Analyzers (ABI PRISM 3730xl, Applied Biosystems, USA). The nucleotide sequences obtained were then used to identify the genes using the Basic Local Alignment Search Tool (BLAST) software system available from National Center Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST/) website.

RESULTS

RNA quality and integrity

RNA with 268/280 >1.9 were obtained from AML samples. Electrophoresis in non-denaturing agarose gel showed two intensive bands representing 28S and 18S rRNA (Figure 1).

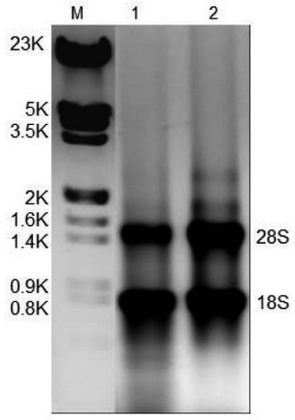

Figure 1: Non-denaturing gel electrophoresis of RNA extracted from AML samples (lane 1 and 2). M: Lambda DNA/ EcoR I + Hind III DNA Marker. *Note:* Due to general differences in form and structure and migration behaviour between single- and double-stranded nucleic acids, 28S rRNA (~5030 nt) and 18S rRNA (~1900 nt) migrate faster through the non-denaturing agarose gel matrix as compared to dsDNA MW marker bands. This resulted in markers positions lower than expected.

The kit recommends that RNA integrity should be checked by gel electrophoresis to observe two bright bands representing the 28S and 18S with ratio of 1.5-2.5:1. A 2:1 ratio for 28S:18S has long been considered the benchmark for intact RNA. However, with the exception of RNA prepared from cultured cells, it is rare to achieve ratio of 2 or greater. Some 28S rRNA contain an AU-rich sequence called a "hidden break" that leads to processing into two smaller RNAs. Even as this piece degrades, the 18S rRNA peak remains fairly constant suggesting this is not associated with large-scale degradation of the RNA sample (TechNotes from ThermoFisher website (https://www.thermofisher.com/my/en/home/references/ambion-tech-support/rna-isolation/tech-notes/assessing-rna-quality.html).

First-strand cDNA synthesis

The cDNA synthesis kit that was used provided a novel, PCR-based method for producing high-quality cDNA from nanogram quantities of total RNA. Unlike commonly used methods, it is able to preferentially enrich for full length cDNAs. A modified oligo(dT)primer (the 3' SMART CDS Primer IIA) primes the first-strand synthesis reaction. When reverse transcription (RT) reaches the 5' end of the mRNA, the enzyme's terminal transferase activity adds a few additional nucleotides, primarily deoxycytidine, to the 3' end of the cDNA. A SMART Oligonucleotide which has an oligo(G) sequence at its 3' end, base-pairs with the deoxycytidine stretch, creating an extended template. RT then switches templates and continues replicating to the end of the oligonucleotide (Super SMART PCR cDNA synthesis kit user manual).

cDNA amplification

After first strand cDNA synthesis, amplification by LD PCR at 15 cycles and every 3 subsequent cycle, is shown in Figure 2. In general, cDNA synthesized from mammalian total RNA appear as a moderately strong smear from 0.5-6 kb with some distinct bands. The optimal number of PCR cycles were determined from the gel before proceeding further. Optimal cycle is one cycle fewer than is needed to reach plateau. Over-cycled cDNA has less representative probes and poor template for cDNA subtraction. As the intensity of the cDNA smear in all tested cycles for the two samples was similarly high, vials amplified to cycle 15 was chosen for further experiment.

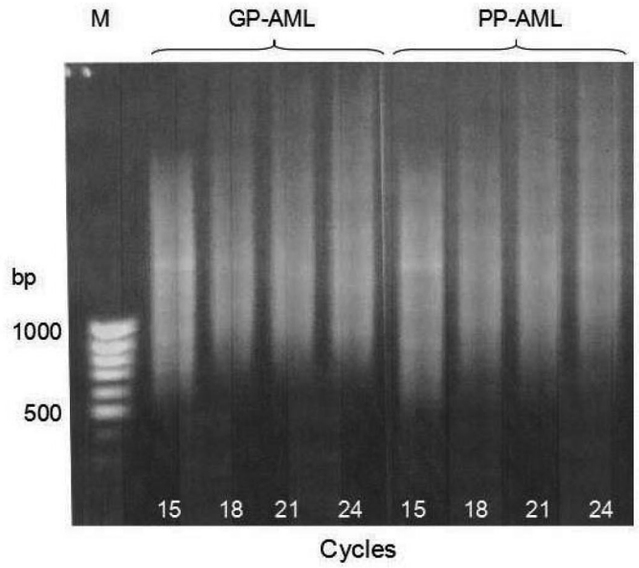

Figure 2: Gel electrophoresis of cDNA amplified at various number of PCR cycles (15, 18, 21, 24). After 15 cycles of amplification, an aliquot of the product was continued for the additional number of cycles. This is to select for product at the exponential phase of amplification. In this experiment, PCR amplification was successfully achieved at 15 cycles. M: 100 bp DNA marker

Column chromatography and Rsa I digestion

Amplified ds cDNA of AML samples purified by column chromatography and subjected to *Rsa I* digestion is shown in Figure 3. Gel analysis after column chromatography was to estimate the PCR products that remained. The yield is typically 50%. The first and second elution was performed to check which had a higher percentage of PCR product. Since the first fraction had a higher yield, this fraction was used for *Rsa I* digestion. *Rsa I* digestion generates shorter, blunt-ended ds cDNA fragments, which are necessary for both adapter ligation and subtraction. After digestion, a smear should range from 0.1-2 kb as shown in the Figure 3, lanes 3.

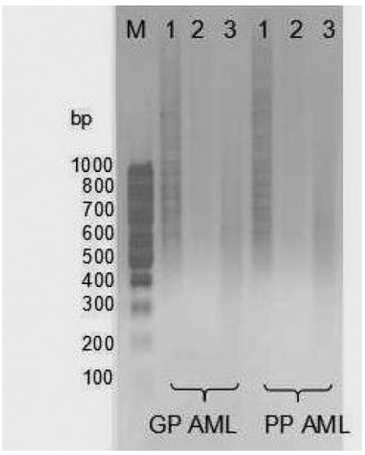

Figure 3: Gel electrophoresis of uncut ds cDNA after LD PCR amplification and PCR purification at first column elution (lanes 1) and second column elution (lanes 2) followed by *Rsa* I digestion (lanes 3) for samples GP-AML and PP-AML, respectively. Before digestion, ds cDNA appears as smear from 0.5 -10kb with bright bands corresponding to abundant mRNA. After digestion smear range from 0.1-0.2 kb. M: 100 bp DNA marker.

Adapter ligation

Adapter ligation efficiency analysis on gel is shown in Figure 4. This PCR experiment to detect G3PDH gene was performed to verify that at least 25% of the cDNAs have adaptors on both ends. Ligation to Adapter 1 was clearly successful for both samples. Ligation to adapter 2R was observed for the PP AML sample though lane 4 was weak. The G3PDH expression varies among tissues. The band intensity for the PCR products should differ no more than four-fold to indicate 25% ligation. For the GP AML sample, lane 4 appeared over-amplified when PCR cycle was increased to 25 cycles. As most bands were present, it was confirmed that a successful ligation experiment was conducted. Also, previous analysis of similar ligation experiments with other samples also showed certain missing/weak bands, however, final results of subtraction were successful (based on subtracted/unsubtracted PCR analysis similar to below). Thus, the missing amplified bands in the efficiency analysis here may be due to missing/poor quality reagents in certain PCR tubes used for this analysis. The original sample was proceeded to the subtractive hybridization steps.

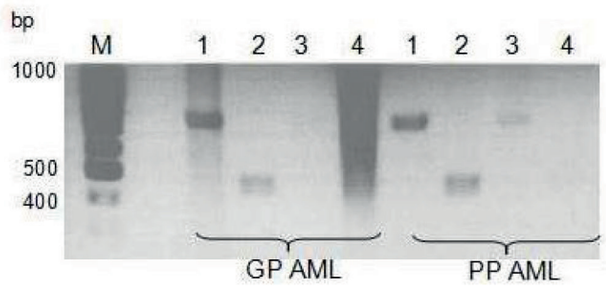

Figure 4: Gel electrophoresis for ligation efficiency analysis. Lane 1: PCR products using Adaptor-1 ligated cDNA as template and the G3PDH 3'Primer and PCR Primer 1. Lane 2: PCR products using Adaptor-1 ligated cDNA as the template and the G3PDH 3' and 5' Primers. Lane 3: PCR products using Adaptor 2R-ligated cDNA as the template and the G3PDH 3; Primers and PCR Primer 1. Lane 4: PCR products using Adaptor 2R-ligated cDNA as the template and the G3PDH 3' and 5' Primers. PCR Primer 1 binds the adapters. Expected band sizes were observed in most lanes. M: 100 bp DNA marker.

Subtractive hybridization

PCR analysis of subtraction for GP-AML and PP-AML samples are shown in Figure 5. Subtracted lanes showed amplification of certain bands distinct from original samples suggestive of successful subtraction.

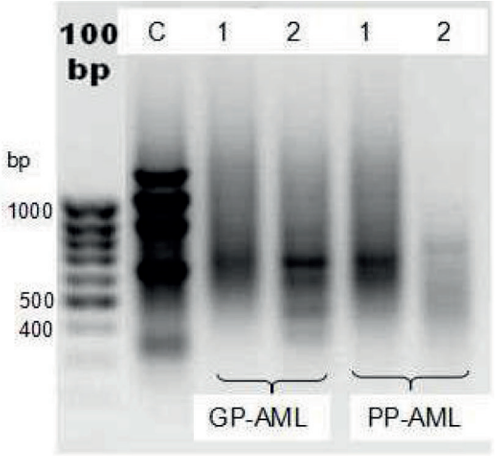

Figure 5: Gel electrophoresis of unsubtracted (lanes 1) and hybridization subtracted (lanes 2) PCR product after nested PCR amplification of the GP and PP AML samples. C is PCR control subtracted cDNA. 100 bp DNA ladder was used as size marker.

Cloning and colony PCR amplification

Colony PCR performed on colonies selected from agar plates are shown in Figure 6.

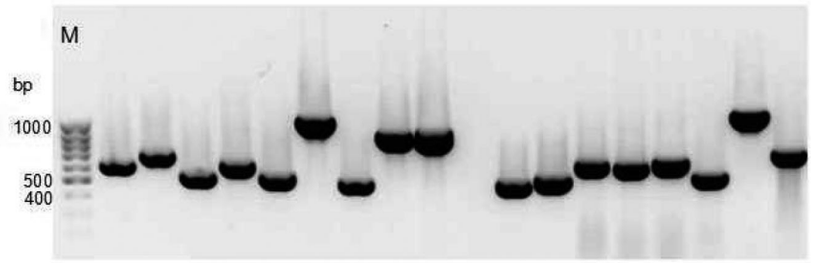

Figure 6: Representative gel electrophoresis of colony PCR products from individual colonies randomly selected from LB-ampicillin agar plates. Each lane shows PCR product from a single colony amplified using the *pGEM-T primers*. 100 bp DNA ladder was used as size marker.

A total of 93 and 91 clones were selected from subtracted libraries of GP-AML and PP-AML, respectively. These consisted of the majority of colonies that grew in the agar plates. Interesting genes however, may still been missed with this method.

Distinct genetic profiles in differentially expressed genes from subtracted library of GP-AML and PP-AML samples

Single pass DNA sequencing of the PCR products provided DNA sequences of about 400 bp. These were submitted into the BLAST software from NBI website which identified the specific gene subtracted. List of identified genes from the GP- and PP-AML samples are shown in Table 1 and Table 2. Genes were grouped based on known functions provided from online gene ontology programs and databases (Nature Protocols; HGNC database). Gene functions were also obtained directly from literature. A large number of the transcripts in these libraries could not be identified and were listed as chromosome contigs. The tables also show the number of clones picked up that were later identified to have the same gene.

Table 1: Identity and characteristics of differentially expressed genes and number of clones isolated from subtracted library of AML samples with good prognosis (GP)

Clone/gene (GP)	Location	Ref	Clones (No)	Clones/genes (PP)	Location	Ref	Clones (No)
Oxidative Phosphorylation							
COX2	Mito	[8]	11	Actin Cytoskeleton	12q13.1	[14]	6
COX3	Mito	[8]	3	*MALAT1*(NR_002819.1)	7p22	[8]	2
ATP8	Mito	[8]	7	*ACTB*(NM_001101.3)	3p21.3	[8]	1
ATP6	Mito	[8]	7	*RHOA*(NM_001664.2)	12p12.3	[8]	2
MTND4	Mito	[8]	3	*ARHGDIB*(NM_001175.4)	12q13.13	[8]	1
ATP5B(NM_001686.3)	12p13.3	[8]	1	*MYL6*(NM_021019.3)	Xp22.2	[8]	1
COX7B(NM_001866.2)	Xq21.1	[8]	1	*TMSB4X*(NM_021109.3)	2p21.3-p21.1	[8]	1
C16orf61/	16q23.2	[9]	1	*CALM2*(NM_001743.3)			
CMC2(NM_020188.2)	Mito	[10]	3				
MTRNR2(16S rRNA)							
Iron Binding Protein				HLA genes			
FTL(NM_000146.3)	19q13.33	[8]	1	*HLA-DRA*(NM_019111.3)	6p21.3	[9]	5
				B2M (NM_004048.2)	15q21-q22.2	[9]	1

Ribosome Biogenesis /Protein Synthesis				Cell Death Regulator			
NOLA2(NM_017838.3)	5q35.3	[11]	1	*CSTB*(NM_000100.2)	21q22.3	[8]	1
RPLP0(NM_001002.3)	12q24.2	[8]	3				
RPL9 (NM_000661.4)	4p13	[8]	1	Stress response/maintenance of mitochondria			
RPL21(NM_000982.3)	13q12.2	[8]	1	*HSPA9*(NM_004134.5)	5q31.1	[8]	1
RPL27(NM_000988.3)	17p21	[8]	2				
RPL27A(NM_000990.4)	11p15	[8]	1	Enzyme			
RPL36A(NM_021029.4)	Xq22.1	[8]	1	*MPO*(NM_000250.1)	17q23.1-q23	[8]	1
RPL41(NM_021104.1)	12q13	[8]	2				
EEF1A1(NM_001402.5)	6q14.1	[8]	2	*SPH*(NM_004577.3)	7p11.2	[8]	1
EIF4A1(NM_001416.2)	17p13	[8]	2				
PCID1/EIF3(NM_006360.3)	11p13	[12]	1				
C14orf166(NM_016039.2)	14q22.1	[9]	1				
Chromatin Remodeling				Unknown			
H2AFZ(NM_002106.3)	4q23	[9] [8]	5	*C6orf115/ABRAL* (XM_941139.2)	6q24.1	[8]	1
PTMA(NM_002823.4)	2q37.1	[13]	1				
OGT(NM_181672.1)	Xq13		1	Chromosome contig			3
Total clones							91

Table 2: Identity and characteristics of differentially expressed genes and number of clones isolated from subtracted library of AML samples with poor prognosis (PP)

Clone/gene (GP)	Location	Ref	Clones (No)	Clones/genes (PP)	Location	Ref	Clones (No)
Aerobic glycolysis/Hypoxia				Hematopoiesis			
PGK1(NM_000291.3)	Xq13.3	[9]	1	*SELL*(NM_000655.3)	1q23-q25	[25]	11
MAPK8(NM_139046.1)*	10q11	[15]	1	*RHOA*(NM_001664.2)	3p21.3	[8]	1
PIM1(NM_002648.)	6p21	[16]	1	*IL3RA*(NM_002183.2)	Xp22.3 & Yp13.3	[26]	1
HIGD1A(NM_014056.3)	3p22.1	[9]	1				
GLUL(NM_002065.4)	1q31	[17]	1	Chemokine			
				CCI1(NM_002981.1)	17	[27]	4
Oxidative Phosphorylation				HOX-related			
COX2	Mito	[8]	1	*HOXA3*(NM_153632.2)	7p15.2	[28]	1
MTND4	Mito	[8]	1	*PBX3*(NM_006195.5)	9q33.3	[28]	2
MTND5	Mito	[8]	1	*ENPP4*(NM_014936.4)	6p12.3	[29]	1
				HOX silencing			
				SF3B1(NM_012433.2)	2q33.1	[30]	2
				PHF2(NM_005392.3)	9q22	[31]	1
				PPIE(NM_203456.1)	1p32	[32]	1
Iron Binding Protein				Cell Death Regulators			
HBD(NM_000519.3)	11p15.5	[9]	1	*PDCD6IP*(NM_013374.3)	3p22.3	[8]	1
				SON(NM_032195.1)	21q22.1-q22.2	[8]	1
Glucocorticoid signaling				*DDB2*(NM_000107.2)	11p12-p11	[8]	1
GLCCI1	7p22.2	[18]	1	*BCAP31*(NM_005745.6)	Xq28	[33]	1
Associated with cancer and neurological disease							
NBPF15(NM_173638.2)	1q36.13	[19]	1				

Ribosome Biogenesis /Protein Synthesis				Methyltransferase			
MRPS21(NM_031901.4)	1q21	[8]	1	*HNMT*(NM_006895.2)	2q22.1	[9]	2
RPS4Y1(NM_001008.3)	Yp11.3	[8]	1	*TPMT*(NM_000367.2)	6p22.3	[8]	1
RPL19(NM_000981.3)	17q12	[8]	1	Calcium Transporter			
BMS1P5 (NM_003611.1)	10q11.22	[9]	1	*ATP2B1*(NM_001682.2)	12q21.33	[9]	1
				Unknown			
Cell Division				*TTTY15*(NR_001545.2)	Yq11.1	[9]	1
KLHDC5(NM_020782.1)	12p11.22	[9]	1				
ANAPC5(NM_016237.3)	12q24	[9]	1				
Fetal/Embryonic				Chromosome contig			37
LCOR(NM_032440.2)	10q24	[20]	1				
CNOT1(NM_016284.3)	16q21	[21]	1				
ORMDL1(NM_016467.4)	2q32	[22]	1				
C*ENPC1*(NM_001812.2)	4q13.2	[23]	1				
NRG4(NM_138573.2)	15q23	[24]	1				
Total clones							93

#MAPK8 gene was isolated fused to 16S RNA

DISCUSSION

In this study, selection of samples was based on extreme ends in prognosis of patient and i*n vitro* viability of cells. Other than elderly (of which neither of these samples were), gender is not a prognostic factor in AML (5).

Gentles et al. (2010) profiled leukaemia stem cell (LSC)-samples enriched from primary AML (N=7) of and normal patient samples (N=7) using global gene expression (microarray). High scores from expression levels of 52 LSC-related genes on 1047 patients were associated with worst relapse free survival (2). Other than the homeobox gene, no other genes were similar to those reported here. This may be due to the different groups of samples used in this study. Also, genes in this study could not be compared as analysis on their function was not presented. Pooling samples data does not allow the phenotype of individual samples to be examined, however, it is a good way to identify prognostic markers. Validation of the microarray identified genes further reduces the number suitable for clinical use (4).

This study differs from others in attempting to study the biological properties within individual acute myeloid leukaemia (AML) samples of clinical significance. Profiling leukemic blasts may provide clues to understanding response to treatment and defining characteristics of the normal counterpart. The importance of these genes as prognostic markers will require validation on a larger number of patient samples.

Here, expression of early developmental stages including embryonic, fetal (*LCOR, CNOT1, ORMDL1, CENPC1, NRG4)* and *HOX*-related genes (*HOXA3, PBX3, ENPP4, SF3B1, PHF2, PPIE*) was seen in the poor prognosis (PP) sample. Transcriptional programs related to embryonic rather than adult stem cells were observed in leukemia stem cells (LSC)(34) and associated with poor survival (35).

Early relapse may be due to dormant immature cells characterized by higher frequency of CD34+(CD38-) (35). A higher percentage of CD34+ cells was also found in the poor prognosis (PP) sample. The differential level of expression compared to the good prognosis sample was unclear since no clones on CD34 transcript was identified. Nevertheless, CD34 binds to L-selectin (36). *SELL* (or *L-selectin*) was the highest clone identified in this study. Selectins promote tumor metastasis by stimulating other molecules such as integrins, chemokines and a permissive microevironment (reviewed in Laubli and Borsig, 2010) (37). *SELL, CCL1* and interleukin (*IL)-3R* observed in the PP sample may be involved in this role. Others also reported the interleukin-3 alpha chain receptor, CD123 expression to be specific to leukaemia stem cells and predictive of an adverse outcome in AML (38).

Activation of IL-3R leads to downstream initiation of the glycolytic pathway to generate energy and ATP and at the same time suppresses oxidative phosphorylation leading to aerobic glycolysis (39). Human pluripotent cells rely mostly on glycolysis to meet their energy demands (40). Similar dependence of certain cancer cells on glycolysis was first reported by Warburg 1956 and referred to as the Warburg effect. Higher expression of *PGK1* of the glycolytic pathway and other IL-3 mediators in aerobic glycolysis such as *MAPK8/JNK1, GLUL(*41) and *PIM1*, an oncogene

expressed in a wide range of tumors including of hematopoietic origin supports early origin of this leukaemic blast. Association of *GLUL* with shorter disease-free survival in cancer patients has been reported (42).

HOX-related genes such as *HOXA3, PBX3, ENPP4* and *HOX* silencing genes such as *SF3B1, PHF2, PPIE w*ere genes expressed higher in the PP sample. *HOX* is a family of genes expressed early in life and function as master transcriptional regulators that control various aspects of morphogenesis and cell development. Other than a distinctive homeodomain, *HOX* are expressed in a spatio-temporal fashion. Four major clusters (*HOXA-D*) encode at least 39 class I genes on four different chromosomes (28). *HOX* genes are found in normal adult hematopoietic stem cells and normal hematopoiesis in differentiation and proliferation with *HOX* cofactors. *HOXA3*, expressed in embryos prevents hematopoietic differentiation. Distinct patterns of *HOX* expression in AML are already reported (22).

SF3B1 is known to silence expression of *HOX* genes by forming complexes with proteins of the Polycomb group (PcG). SF3B1 and PcG proteins colocalize on the 5'regions of individual *HOX* genes (30). *HOXA3* is located towards the 3' region. *PPIE*, a nuclear cyclophilin binds to *MLL* and in leukemia cells has been shown to alter expressions of *HOX* genes which are targets of *MLL* regulation (32). Rearrangements involving *MLL* are of clinical significance in AML (5).

LCoR and *CNOT1* are ligand dependent co-repressors (43) highly expressed in embryonic cells. *LCoR* binds directly to specific *HDAC*s (histone deacetylases) which is associated with gene silencing and altered expression and mutations of genes that encode *HDAC*s have been linked to tumor development (44).

In contrast, the good prognosis sample exhibited increased activity utilizing mitochondrial oxidative phosphorylation for production of energy and expressing many genes associated with this pathway. This pathway is regulated by protein complexes known as the respiratory chain encoded by two genetic systems: the mitochondrial DNA (mtDNA) and nuclear DNA (nDNA). The mtDNA codes for 13 proteins including 7 NADH dehydrogenase (ND) subunits for complex I, a cytochrome b for complex III, 3 cytochrome c oxidase (COX) subunits for complex IV, and 2 ATPase (ATPase6/8) for complex V. Complex II is solely encoded by nDNA (45). The adverse contribution of altered cytochrome c oxidase subunits in AML has been reported (46).

Approximately one third of the genes identified in the GP sample were involved in ribosome biogenesis or protein synthesis which is increased in cells undergoing proliferation. Transcription includes steps in initiation, elongation and termination (47).

Ribosomes are made up of half protein and half RNA. Ribosomal RNA synthesis occurs in the nucleolus which is then assembled with 82 ribosomal proteins to become ribosomes (47) were well represented in the GP library.

Chromatin remodeling and actin cytoskeleton genes are elements that serve in maintenance of cellular shape as well as cellular migration, cell division and cell remodeling. *TMSB4X* encodes an actin sequestering protein involved in angiogenesis, cell proliferation, migration, and differentiation (48). Studies show cell migration indicates poor prognosis (14).

CONCLUSION

Thus, subtractive hybridization isolated distinct gene patterns from the good and poor prognosis AML patient samples. Gene functions as identified in Table 1 and Table 2 were studies of normal stem cells and thus patterns here were reflective of normal hematopoietic hierarchy. The poor prognosis sample contained many stem cell associated genes suggestive of its leukemic stem cell origin. Many of the genes were also cancer related and are potential prognostic markers and targets for therapy.

ACKNOWLEDGEMENT

This study was supported by research grants from the Ministry of Science, Technology and Innovation (Project No: 02-01-04-SF008) and MAKNA Research Award, 2003. We would like to thank the Department of Haematology and Department of Paediatric, Universiti Kebangsaan Malaysia Medical Centre for their support of this project.

REFERENCES

1. Juliusson G, Antunovic P, Derolf A, Lehmann S, Möllgård L, Stockelberg D, *et al*. Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. Blood 2009; 113(18):4179-87.

2. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. JAMA. 2010; 304(24):2706-15.

3. De Jonge HJ, Huls G, de Bont ES. Gene expression profiling in acute myeloid leukaemia. Neth J Med. 2011; 69(4):167-76.

4. Tian Y, Huang Z, Wang Z, Yin C, Zhou L, Zhang L, *et al*. Identification of novel molecular markers for prognosis estimation of acute myeloid leukemia: over-expression of PDCD7, FIS1 and Ang2 may indicate poor prognosis in pretreatment patients with acute myeloid leukemia. PLoS One. 2014; 9(1):e84150.

5. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, *et al*. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood 2009; 114(5):937-51.

6. Maha A, Cheong SK, Leong CF, Seow HF. Cell viability of acute myeloid blasts in culture correlates with treatment outcome. Hematology 2008; 13(1): 13-20.

7. Bruserud O, Gjertsen BT, Foss B, Huang TS. New strategies in the treatment of acute myelogenous leukemia (AML): in vitro culture of aml cells--the present use in experimental studies and the possible importance for future therapeutic approaches. Stem Cells. 2001; 19(1):1-11.

8. Nature Protocols. 2009; 4(1):44 & Genome Biology. 2003; 4(5):P3.

9. HGNC Database, HUGO Gene Nomenclature Committee (HGNC), EMBL Outstation - Hinxton, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK www.genenames.org. Jun 2013

10. Burk A, Douzery EJP, Springer MS. The secondary structure of mammalian mitochondrial 16S rRNA molecules: refinements based on a comparative phylogenetic approach. J Mamm Evol 2002; 9(3): 225-52.

11. Tollervey D, Kiss T. Function and synthesis of small nucleolar RNAs. Curr Opin Cell Biol 1997; 9: 337-42.

12. Goh SH, Hong SH, Hong SH, Lee BC, Ju MH, Jeong JS, *et al*. eIF3m expression influences the regulation of tumorigenesis-related genes in human colon cancer. Oncogene 2011; 30(4): 398-409.

13. Sakabe K, Hart GW. O-GlcNActransferase regulates mitotic chromatin dynamics. J Biol Chem 2010; 285(45): 34460-68.

14. Schmidt LH, Spieker T, Koschmieder S, Schäffers S, Humberg J, Jungen D, *et al*. The long noncoding MALAT-1 RNA indicates a poor prognosis in non-small cell lung cancer and induces migration and tumor growth. J Thorac Oncol 2011; 6(12): 1984-92.

15. Deng H, Yu F, Chen J, Zhao Y, Xiang J, Lin A**.** Phosphorylation of Bad at Thr-201 by JNK1 promotes glycolysis through activation of phosphofructokinase-1. Biol Chem 2008; 283(30): 20754-60.

16. Chen J, Kobayashi M, Darmanin S, Qiao Y, Gully C, Zhao R, *et al*. Hypoxia-mediated up-regulation of *Pim-1* contributes to solid tumor formation. Am J Pathol.2009; 175(1): 400-11.

17. DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, *et al*. Beyond aerobic glycolysis: Transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. Proc Natl Acad Sci 2007; 104: 19345–50.

18. Galloway JL, Wingert RA, Thisse C, Thisse B, Zon LI. Combinatorial regulation of novel erythroid gene expression in zebrafish. Exp Hematol 2008; 36(4): 424–32.

19. Vandepoele K, Van Roy N, Staes K, Speleman F, van Roy F. A novel gene family NBPF: intricate structure generated by gene duplications during primate evolution. Mol Biol Evol 2005; 22(11): 2265-74.

20. Fernandes I, Bastien Y, Wai T, Nygard K, Lin R, Cormier O. Ligand-dependent nuclear receptor corepressor LCoR functions by histone deacetylase-dependent and -independent mechanisms. Mol Cell 2003; 11(1):139-50.

21. Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK. Cnot1, Cnot2, and Cnot3 maintain mouse and human ESC identity and inhibit extraembryonic differentiation. Stem Cells 2012; 30(5):910-22.

22. Hjelmqvist L, Tuson M, Marfany G, Herrero E, Balcells S, Gonzàlez-Duarte R. ORMDL proteins are a conserved new family of endoplasmic reticulum membrane proteins. Genome Biol 2002 3(6):RESEARCH0027.

23. Kalitsis P, Fowler KJ, Earle E, Hill J, Choo KH.. Targeted disruption of mouse centromere protein C gene leads to mitotic disarray and early embryo death. Proc Natl Acad Sci USA. 1998; 95(3): 1136–1141.

24. Huotari MA, Miettinen PJ, Palgi J, Koivisto T, Ustinov J, Harari D, *et al*. ErbB signaling regulates lineage determination of developing pancreatic islet cells in embryonic organ culture. Endocrinology. 2002; 143(11):4437-46.

25. Baumhueter S, Dybdal N, Kyle C, Lasky LA. Global vascular expression of murine CD34, a sialomucin-like endothelial ligand for L-selectin. Blood. 1994; 84(8): 2554-65.

26. Mroczko B, M Szmitkowski. Hematopoietic cytokines as tumor markers. Clin Chem Lab Med. 2004; 42(12): 1347-54.

27. Spinetti G, Bernardini G, Camarda G, Mangoni A, Santoni A, Capogrossi MC, *et al*. The chemokine receptor CCR8 mediates rescue from dexamethasone-induced apoptosis via an ERK-dependent pathway. J Leukoc Biol 2003; 73(1): 201-207.

28. Drabkin HA, Parsy C, Ferguson K, Guilhot F, Lacotte L, Roy L, *et al*. Quantitative HOX expression in chromosomally defined subsets of acute myelogenous leukemia. Leukemia 2002; 16(2): 186-95.

29. Massé K, Bhamra S, Allsop G, Dale N, Jones EA. Ectophosphodiesterase/nucleotide phosphohydrolase (Enpp) nucleotidases: cloning, conservation and developmental restriction. Int J Dev Biol 2010; 54(1):181-93.

30. Isono K, Mizutani-Koseki Y, Komori T, Schmidt-Zachmann MS, Koseki H. Mammalian polycomb-mediated repression of Hox genes requires the essential spliceosomal protein Sf3b1. Genes Dev 2005; 19(5): 536-41.

31. Wang J, Du G, Zhou Y, Yin B, Yuan J, Qiang B. Cloning and identification of PHF2 cDNA and its alternatively spliced transcripts. Chinese Science Bulletin. 1999; 44 (15):1382-87.

32. Fair K, Anderson M, Bulanova E, Mi H, Tropschug M, Diaz MO. Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. Mol Cell Biol 2001; 21(10): 3589-97.

33. Breckenridge DG, Stojanovic M, Marcellus RC, Shore GC**.** Caspase cleavage product of BAP31 induces mitochondrial fission through endoplasmic reticulum calcium signals, enhancing cytochrome c release to the cytosol. J Cell Biol 2003; 160(7): 1115-27.

34. Somervaille TC, Matheny CJ, Spencer GJ, Iwasaki M, Rinn JL, Witten DM, *et al*. Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. Cell Stem Cell 2006; 4(2):129-40.

35. Van Rhenen A, Feller N, Kelder A, Westra AH, Rombouts E, Zweegman S, *et al*. High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. Clin Cancer Res 2005; 11(18):6520-27.

36. Baumheter S, Singer MS, Henzel W, Hemmerich S, Renz M, Rosen SD, *et al*. Binding of L-selectin to the vascular sialomucin CD34. Science 1993; 262(5132):436-38.

37. Läubli H, Borsig L. Selectins promote tumor metastasis. Semin Cancer Biol. 2010; 20(3):169-77.

38. Vergez F, Green AS, Tamburini J, Sarry JE, Gaillard B, Cornillet-Lefebvre P, *et al*. High levels of CD34+CD38low/-CD123+ blasts are predictive of an adverse outcome in acute myeloid leukemia: a Groupe Ouest-Est des Leucemies Aigues et Maladies du Sang (GOELAMS) study. Haematologica. 2011; 96(12):1792-8.

39. Bauer DE, Harris MH, Plas DR, Lum JJ, Hammerman PS, Rathmell JC, *et al*. Cytokine stimulation of aerobic glycolysis in hematopoietic cells exceeds proliferative demand. FASEB J 2004; 18(11):1303-5.

40. Varum S, Rodrigues AS, Moura MB, Momcilovic O, Easley CA 4th, Ramalho-Santos J, *et al*. Energy metabolism in human pluripotent stem cells and their differentiated counterparts. PLOS One 2011; 6(6): e20914.

41. Vermeulen T, Görg B, Vogl T, Wolf M, Varga G, Toutain A, *et al*. Glutamine synthetase is essential for proliferation of fetal skin fibroblasts. Arch Biochem Biophys 2008; 478(1): 96-102.

42. Yang L, Moss T, Mangala LS, Marini J, Zhao H, Wahlig S, *et al*. Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. Mol Syst Biol 2014; 10:728.

43. Winkler GS, Mulder KW, Bardwell VJ, Kalkhoven E, Timmers HT. Human Ccr4-Not complex is ligand-dependent repressor of nuclear receptor-mediated transcription. EMBO J 2006; 25(13): 3089-99.

44. Ropero S, Esteller M. The role of histone deacetylases (HDACs) in human cancer. Mol Oncol 2007; 1: 19–25.

45. Carew JS, Huang P. Mitochondrial defects in cancer. Mol Cancer 2002; 9: 1-9.

46. Silkjaer T, Nyvold CG, Juhl-Christensen C, Hokland P, Nørgaard JM. Mitochondrial cytochrome c oxidase subunit II variations predict adverse prognosis in cytogenetically normal acute myeloid . Eur J Haematol 2013; 91(4):295-303.

47. Snustad DP, Simmons MJ: Principles of genetics. 3rded. New York: John Wiley & Sons, Inc. 2003.

48. Gauger KJ, Hugh JM, Troester MA, Schneider SS. Down-regulation of sfrp1 in a mammary epithelial cell line promotes the development of a CD44high/CD24low population which is invasive and resistant to anoikis. Cancer Cell Int 2009; 9: 11.